15th Triennial World Congress, Barcelona, Spain

WEAK CONVERGENCE ANALYSIS OF CROSS-COUPLED KALMAN FILTER STATE-ESTIMATION ALGORITHM FOR BILINEAR SYSTEMS

Vladislav B. Tadić [*] **Vikram Krishnamurthy** [**]

[*] *Department of Electrical and Electronic Engineering*
The University of Melbourne, Parkville, Victoria 3010, Australia
e-mail: v.tadic@ee.mu.oz.au
[**] *Department of Electrical and Electronic Engineering*
The University of Melbourne, Parkville, Victoria 3010, Australia
e-mail: vikram@ee.mu.oz.au

Abstract: In this paper, we present an asymptotic analysis of a recursive cross-coupled Kalman filter algorithm for estimating the state of a partially observed bilinear stochastic system. The cross-coupled Kalman filter algorithm consists of two Kalman filters – each Kalman filter estimating the state of one of the two state components of the bilinear system. Our asymptotic analysis provides weak convergence results on the tracking capabilities of the resulting cross-coupled Kalman filter algorithm.

Keywords: State estimation, bilinear systems, Kalman filters,stochastic approximation, weak convergence analysis.

1. INTRODUCTION AND MOTIVATION

Bilinear models (Fnaiech and Ljung, 1987), (Priestley, 1991) are widely used to model nonlinear processes in signal and image processing and communication systems modeling. In particular, they arise in areas such as channel equalization (Benedetto and Biglieri, 1983), nonlinear tracking (Halawani *et al.*, 1984) and many other areas of engineering, socioeconomics and biology (Bruni *et al.*, 1974).

Let $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_0^+ = [0, \infty)$, $\mathbb{R}_0^- = (-\infty, 0]$, $\mathbb{R}^+ = (0, \infty)$ and $\mathbb{R}^- = (-\infty, 0)$. In this paper, for discrete time $n \geq 0$, we consider estimating the states x_n and s_n of the following partially observed scalar bilinear system given the observation sequence $Y_n = (y_0, \dots, y_n)$:

$$x_{n+1} = s_n x_n + u_{n+1}, \tag{1.1}$$

$$y_n = c x_n + v_n, \tag{1.2}$$

$$s_{n+1} = (1 + \gamma\lambda) s_n + \sqrt{\gamma} w_{n+1}. \tag{1.3}$$

$\gamma \in \mathbb{R}^+$ is a small parameter, $\lambda \in \mathbb{R}_0^-$ and $c \in \mathbb{R}$ are constants, while $\{u_n\}_{n \geq 0}$, $\{v_n\}_{n \geq 0}$ and $\{w_n\}_{n \geq 0}$ are sequences of i.i.d. $\mathbb{R}$-valued zero-mean random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Due to the widespread use of the above bilinear model, there is strong motivation to develop estimation algorithms for the states $\{x_n\}_{n\geq 0}$ and $\{s_n\}_{n\geq 0}$ of such systems given noisy observations $\{y_n\}_{n\geq 0}$. Unfortunately, the optimal filter for reconstructing conditional mean estimates $\mathbb{E}(x_n|Y_n)$ and $\mathbb{E}(s_n|Y_n)$ of the partially observed bilinear system cannot be characterized by a finite dimensional statistic. Thus, practical state estimation algorithms for bilinear systems are necessarily suboptimal. For example, the extended Kalman filter (EKF) is an approximate filter that linearizes around conditional mean state estimates at each time instant.

In recent work (Johnston and Krishnamurthy, 1999), rather than computing approximations of the conditional mean estimates, iterative finite dimensional algorithms were presented for computing the optimal MAP state sequence estimate for a bilinear system. In particular, the Expectation Maximization (EM) algorithm (Dempster *et al.*, 1977), (Wu, 1983) was used to numerically compute these MAP state sequence estimates. Somewhat surprisingly, the EM algorithms in (Johnston and Krishnamurthy, 1999) for estimating the state of the bilinear system, involved cross-coupling two Kalman smoothers, each estimating one of the two component signals of the bilinear system.

The aim of this paper is to present and prove weak convergence of a recursive (on-line) version of the EM algorithm for state estimation of bilinear systems. In analogy to the off-line EM algorithm presented in (Johnston and Krishnamurthy, 1999), the recursive (on-line) algorithm we present cross-couples two Kalman filters, each Kalman filter estimating one of the two component signals of the bilinear system. For convenience we call the algorithm as the *cross-coupled Kalman filter algorithm.* Empirical studies on the performance of such cross-coupled filters have shown remarkable improvements compared to the standard Extended Kalman filter (EKF), see (Johnston and Krishnamurthy, 1999) for extensive numerical examples. Our main results, Theorem 1 and Corollary 1, provide weak convergence limits for the tracking errors of the estimates of the states $\{s_n\}_{n\geq 0}$ as $n \to \infty$ and $\gamma \to 0+$ of the cross-coupled Kalman filter algorithm.

Motivation – MAP State Estimation

Let $S_N = (s_0, \ldots, s_N)$, $X_N = (x_0, \ldots, x_N)$ and $Y_N = (y_0, \ldots, y_N)$. Consider now the following MAP sequence estimation problem for S_N and X_N: Compute the MAP sequence estimates

$$(\hat{S}_N, \hat{X}_N)^{MAP} = \arg\max_{S,X} f_{S_N,X_N|Y_N}(S, X|Y_N) \tag{1.4}$$

where $f_{S_N,X_N|Y_N}(\cdot,\cdot|\cdot)$ denotes the joint conditional probability density function of S_N and X_N, conditioned on the measurements Y_N. In (Johnston and Krishnamurthy, 1999), the above optimization was carried out iteratively by using the following coordinate descent algorithm (Luenberger, 1984): Staring with (arbitrary) initial estimates $\hat{S}_N^0$, $\hat{X}_N^0$, compute recursively $\hat{S}_N^i$ and $\hat{X}_N^i$, for iterations $i \geq 1$, as

$$\hat{X}_N^{i+1} = \arg\max_{\hat{X}} \log f_{S_N,X_N,Y_N}(\hat{S}_N^i, \hat{X}, Y_N), \tag{1.5}$$

$$\hat{S}_N^{i+1} = \arg\max_{\hat{S}} \log f_{S_N,X_N,Y_N}(\hat{S}, \hat{X}_N^{i+1}, Y_N), \tag{1.6}$$

where $f_{S_N,X_N,Y_N}(\cdot,\cdot,\cdot)$ is the joint probability density function of S_N, X_N, Y_N.

It is proved in (Johnston and Krishnamurthy, 1999) that subject to some mild regularity conditions, the MAP estimates of S_N, X_N generated by the above algorithm converges to a stationary point in the likelihood surface $f_{S_N,X_N|Y_N}(\cdot,\cdot|Y_N)$. Indeed, the above algorithm is shown in (Johnston and Krishnamurthy, 1999) to be a special instance of the Expectation Maximization (EM) algorithm which is used widely for MAP and maximum likelihood estimation (Dempster *et al.*, 1977).

To motivate our recursive (on-line) algorithm first consider the implementation of (1.5) and (1.6). It is straightforward to show that (1.5) is equivalent to

$$\begin{aligned}\hat{X}_N^{i+1} = \arg\max_{\hat{X}=(\hat{x}_0,\ldots,\hat{x}_N)} \Bigg(&- 2^{-1}\sigma_{x_0}^{-2}(\hat{x}_0 - \bar{x}_0)^2 \\ &+ 2^{-1}\sigma_v^{-2}\sum_{n=1}^{N}(y_n - c\hat{x}_n)^2 \\ &- 2^{-1}\sigma_u^{-2}\sum_{n=1}^{N}(\hat{x}_n - \hat{s}_{n-1}^i\hat{x}_{n-1})^2 \\ &+ \text{terms independent of } \hat{X}\Bigg),\end{aligned} \tag{1.7}$$

where $\hat{S}_N^i = (\hat{s}_0^i, \ldots, \hat{s}_N^i)$, σ_{x_0}, σ_u and σ_v are covariances of x_0, u_0 and v_0 (respectively), while $\bar{x}_0$ is the mean of x_0. On the other hand, (1.7) is identically the maximum of the log likelihood of the linear Gaussian signal model:

$$\tilde{x}_{n+1} = \hat{s}_n^i\tilde{x}_n + u_{n+1}, \quad 0 \leq n < N, \tag{1.8}$$

$$\tilde{y}_n = c\tilde{x}_n + v_n, \quad 0 \leq n \leq N. \tag{1.9}$$

Hence, the maximization (1.7) is carried out via a Kalman smoother on the linear signal model (1.8), (1.9). Similarly, it is straightforward to show that (1.6) is equivalent to

$$\begin{aligned}\hat{X}_N^{i+1} = \arg\max_{\hat{S}=(\hat{s}_0,\ldots,\hat{s}_N)} \Bigg(&- 2^{-1}\sigma_{s_0}^{-2}(\hat{s}_0 - \bar{s}_0)^2 \\ &+ 2^{-1}\sigma_u^{-2}\sum_{n=1}^{N-1}(\hat{x}_{n+1}^{i+1} - \hat{x}_n^{i+1}\hat{s}_n)^2 \\ &- 2^{-1}\gamma^{-2}\sigma_w^{-2}\sum_{n=1}^{N}(\hat{s}_n - (1+\gamma\lambda)\hat{s}_{n-1})^2 \\ &+ \text{terms independent of } \hat{S}\Bigg),\end{aligned} \tag{1.10}$$

where $\hat{X}_N^{i+1} = (\hat{x}_0^{i+1}, \ldots, \hat{x}_N^{i+1})$, σ_{s_0} and σ_w are covariances of s_0, and w_0 (respectively), while $\bar{s}_0$

is the mean of s_0. On the other hand, (1.10) is identically the maximum of the log likelihood of the linear Gaussian signal model:

$$\tilde{s}_{n+1} = (1+\gamma\lambda)\tilde{s}_n + \gamma w_{n+1}, \quad 0 \le n < N, \quad (1.11)$$

$$\tilde{y}_n = \tilde{c}_n \tilde{s}_n + u_n, \quad 0 \le n < N, \quad (1.12)$$

where $\tilde{y}_n = \hat{x}^{i+1}_{n+1}$ and $\tilde{c}_n = \hat{x}^i_n$, $0 \le n < N$. Hence, the maximization (1.10) is achieved by a Kalman smoother applied to the linear signal model (1.11), (1.12). Thus, the algorithm involves cross-coupling two Kalman smoothers.

The Cross-Coupled Kalman Filter Algorithm for Recursive State Estimation

The above iterative coordinate descent algorithm for computing the MAP state estimates is off-line. Its structure based on the cross–coupling of two Kalman smoothers point towards the following recursive (on-line) algorithm for state estimation: Replace the Kalman smoothers by Kalman filters. The resulting recursive algorithm has a clear intuitive interpretation — knowledge of the signal states $\{s_n\}_{n\ge0}$ results in a Kalman filter achieving optimal estimates of $\{x_n\}_{n\ge0}$. Conversely, knowledge of the states $\{x_n\}_{n\ge0}$ means a Kalman filter achieves optimal estimates of $\{s_n\}_{n\ge0}$. Hence, from a heuristic point of view it makes sense to cross–couple two Kalman filters. In extensive numerical studies presented in (Johnston and Krishnamurthy, 1999), it has been shown that the cross-coupled Kalman filter algorithm performs significantly better than the Extended Kalman filter. The aim of this paper is to study the convergence and tracking properties of the cross-coupled Kalman filter algorithm.

Analysis Limitations

In this paper, the cross-coupled Kalman filter algorithm for state estimation in scalar bilinear systems is considered only. The main reason for not analyzing the multidimensional case comes out from the fact that the 'averaged' ordinary differential equation (ODE) associated with the algorithm and the system (which is defined in (4.2), Section 4) is so complex in the multidimensional case that it is extremely hard to analyze its stability properties. However, without demonstrating that the 'averaged' ODE is globally Lagrange stable, it is practically impossible to provide any result on the asymptotic behavior of the cross-coupled Kalman filter algorithm. Moreover, the properties of the parameterized Markov chain associated with the algorithm and the system (which is defined in (4.1), Section 4) are also very complex and it is extremely difficult to analyze its stability (i.e., ergodicity). However, the properties of the 'averaged' ODE are tightly connected with this Markov chain and without showing its geometric ergodicity, it is not even possible to verify if the right-hand side of the 'averaged' ODE is well-defined.

2. BILINEAR SYSTEM AND CROSS-COUPLED KALMAN FILTERING ALGORITHM

The on-line estimation of s_n given the observations $y_0, \ldots, y_n$ is the problem considered in this paper. As (1.3) is a system with slow dynamics for the case of $\gamma \to 0+$, estimating s_n reduces to the problem of tracking a slowly varying parameter (for details see (Benveniste *et al.*, 1990, Section 4, Part I)). The problem of the estimation of $\{s_n\}_{n\ge0}$ is analyzed under the following assumptions:

A1. $\{u_n\}_{n\ge0}$, $\{v_n\}_{n\ge0}$, $\{w_n\}_{n\ge0}$ are mutually independent. $\mu^8_u = \mathbb{E}|u_0|^8 < \infty$, $\mu^8_v = \mathbb{E}|v_0|^8 < \infty$, $\mu^8_w = \mathbb{E}|w_0|^8 < \infty$, $\mathbb{E}(u_0) = \mathbb{E}(v_0) = \mathbb{E}(w_0) = 0$, $\sigma^2_u = \mathbb{E}|u_0|^2 > 0$, $\sigma^2_v = \mathbb{E}|v_0|^2 > 0$, $\sigma^2_w = \mathbb{E}|w_0|^2 > 0$ and $c \ne 0$.

Assumption A1 corresponds with the noise $\{u_n\}_{n\ge0}$, $\{v_n\}_{n\ge0}$, $\{w_n\}_{n\ge0}$, and is typical for the problems of state-estimation in stochastic systems (for details see (Anderson and Moore, 1979), (Caines, 1988)). It is satisfied if $\{u_n\}_{n\ge0}$, $\{v_n\}_{n\ge0}$, $\{w_n\}_{n\ge0}$ are jointly independent Gaussian white noise.

Remark. In (Tadić and Krishnamurthy, 2001), a mean-square trarking error of the cross-coupled Kalman filter has been analyzed. The analysis has been carried out for a bilinear system which is slightly more general than (1.1) – (1.3).

Cross Coupled Kalman Filter algorithm: The following algorithm is used for the state estimation of the system (1.1) – (1.3):

$$\hat{s}_{n+1} = P_Q((1+\gamma\lambda)\hat{s}_n + (1+\gamma\lambda)\gamma\hat{q}_n\hat{x}_n(\gamma\hat{q}_n\hat{x}^2_n + \sigma^2_u)^{-1}(\hat{x}_{n+1} - \hat{s}_n\hat{x}_n)), \quad (2.13)$$

$$\hat{q}_{n+1} = P((1+\gamma\lambda)^2\hat{q}_n - (1+\gamma\lambda)^2\gamma\hat{q}^2_n\hat{x}^2_n(\gamma\hat{q}_n\hat{x}^2_n + \sigma^2_u)^{-1} + \gamma\sigma^2_w), \quad (2.14)$$

$$\hat{x}_{n+1} = \hat{s}_n\hat{x}_n + c\hat{p}_{n+1}(c^2\hat{p}_{n+1} + \sigma^2_v)^{-1}(y_{n+1} - c\hat{s}_n\hat{x}_n), \quad (2.15)$$

$$\hat{p}_{n+1} = \hat{s}_n^2\hat{p}_n - c^2\hat{s}_n^2\hat{p}_n^2(c^2\hat{p}_n + \sigma_v^2)^{-1} + \sigma_u^2. \quad (2.16)$$

$\hat{s}_0$, $\hat{x}_0$ are $\mathbb{R}$-valued deterministic variables, while $\hat{p}_0$, $\hat{q}_0$ are $\mathbb{R}_0^+$-valued deterministic variables. $Q = [-\delta, \delta]$, where $\delta \in (0,1)$ is a constant, while the projections are defined as

$$P_Q(\hat{s}) = \arg\min_{\hat{s}' \in Q} |\hat{s} - \hat{s}'|,$$
$$P(\hat{q}) = \arg\min_{\hat{q}' \in [0,k]} |\hat{q} - \hat{q}'|, \quad \hat{s}, \hat{q} \in \mathbb{R}, \quad (2.17)$$

where $k = 2(1 + c^{-2}\sigma_u^{-2}\sigma_v^2)^2(1 + \sigma_w)$.

Notice that (2.13) – (2.16) are the equations of two Kalman filters. The (truncated) Kalman filter (2.13), (2.14) operates on the linear model with state dynamics (1.3) and observation equation (1.1) modified by replacing x_n with its estimate $\hat{x}_n$. The filter (2.13), (2.14) yields a state estimate $\hat{s}_{n+1}$ together with a covariance estimate $\gamma\hat{q}_{n+1}$, while the estimate $\hat{x}_n$ (used by this filter) is computed by the Kalman filter (2.15), (2.16) (notice that this Kalman filter updates the estimates in the predictor form usually denoted as $\hat{s}_{n+1|n}$ and $\hat{q}_{n+1|n}$, i.e., $\hat{s}_{n+1}$ is computed after receiving the observation $\hat{x}_{n+1}$ of s_n). On the other hand, the Kalman filter (2.15), (2.16) operates on the linear model with state dynamics (1.1) and observation equation (1.2) modified by replacing s_n with its estimate $\hat{s}_n$. The filter (2.15), (2.16) yields a state estimate $\hat{x}_n$ together with a covariance estimate $\hat{p}_n$, while the estimate $\hat{s}_n$ (required by this filter) is computed by the Kalman filter (2.13), (2.14) (this Kalman filter operates in the filtered form with estimates usually denoted as $\hat{x}_{n+1|n+1}$ and $\hat{p}_{n+1|n+1}$, i.e., $\hat{x}_{n+1}$ is updated based on the observation y_{n+1} of x_{n+1}). Thus, the two Kalman filters are cross-coupled, each feeding its estimate to the other which in turn computes a new estimate. From a heuristic point of view, if $\hat{x}_n$ is close to x_n, then the Kalman filter (2.13), (2.14) generates near optimal estimates $\hat{s}_{n+1}$. If this estimate is close to the true state s_{n+1}, then the Kalman filter (2.15), (2.16) generates near optimal estimates of $\hat{x}_{n+1}$ and so on.

A natural performance measure for quantifying the performance of the above cross-coupled Kalman filter algorithm is to compare its estimates $\hat{s}_n$ with s_n. Hence, the aim of the analysis carried out in this paper is to provide asymptotic results on the tracking errors $\hat{s}_n - s_n$ and as $n \to \infty$ and $\gamma \to 0+$ of the cross-coupled Kalman filter algorithm.

Remark. In (Ljung, 1979), local stability analysis of the Extended Kalman filter (in the prediction form) is carried out for estimating s_n assuming $s_n = s$ is a constant parameter. It is shown therein that zero is the stationary point of the 'averaged' ODE and that the 'averaged' ODE has three stationary points if $s < 0$. Basically, this means that the identifiability conditions are violated (see (Benveniste *et al.*, 1990, Assumption NS3, Page 124)). Using the filtering form of the Kalman filter (2.15), (2.16) in the cross-coupled Kalman filter algorithm circumvents this problem – for details see (Tadić and Krishnamurthy, 2001).

3. ALGORITHM REPRESENTATION AND NOTATION

In this section, it is shown that the difference equations (1.1) – (1.3) and (2.13) – (2.16) of the partially observed bilinear model together with the cross coupled Kalman filter algorithm falls into the category of stochastic approximation algorithms studied in (Benveniste *et al.*, 1990). For $s, \hat{s}, \hat{q}, x, \hat{x}, \hat{x}' \in \mathbb{R}$, $\theta = (\hat{s}, \hat{q})$, $\xi = (x, \hat{x}, \hat{x}')$, let

$$p(\hat{s}) = 2^{-1}(\sigma_u^2 - c^{-2}\sigma_v^2(1 - \hat{s}^2)) + \left(4^{-1}(\sigma_u^2 - c^{-2}\sigma_v^2(1 - \hat{s}^2))^2 + c^{-2}\sigma_u^2\sigma_v^2\right)^{1/2}, \quad (3.1)$$

$$\alpha(\hat{s}) = \sigma_v^2\hat{s}(c^2p(\hat{s}) + \sigma_v^2)^{-1},$$
$$\beta(\hat{s}) = c^2p(\hat{s})(c^2p(\hat{s}) + \sigma_v^2)^{-1},$$
$$F(\theta, \xi) = \lambda\hat{s} + \sigma_u^{-2}\hat{q}\hat{x}(\hat{x}' - \hat{s}\hat{x}),$$
$$G(\theta, \xi) = \sigma_w^2 + 2\lambda\hat{q} - \sigma_u^{-2}\hat{q}^2\hat{x}^2,$$
$$A(\theta, s) = \begin{bmatrix} s & 0 & 0 \\ 0 & 0 & 1 \\ s\beta(\hat{s}) & 0 & \alpha(\hat{s}) \end{bmatrix},$$
$$B(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \beta(\hat{s}) & c^{-1}\beta(\hat{s}) \end{bmatrix}$$

and $H(\theta, \xi) = (F(\theta, \xi), G(\theta, \xi))$ (notice that $p(\hat{s})$ is the asymptotic solution of the Riccati equation (2.16) parameterized by $\hat{s}$). Moreover, let $\hat{x}'_0 = \hat{x}_0$ and

$$\hat{x}'_{n+1} = \alpha(\hat{s}_n)\hat{x}'_n + \beta(\hat{s}_n)(x_{n+1} + c^{-1}v_{n+1}),$$

$$\begin{aligned}\mu_{n+1} = {} & \sigma_u^{-2}\hat{q}_n(\hat{x}_n\hat{x}_{n+1} - \hat{x}'_n\hat{x}'_{n+1}) \\ & -\sigma_u^{-2}\hat{s}_n\hat{q}_n(\hat{x}_n^2 - |\hat{x}'_n|^2) \\ & +\gamma\lambda\sigma_u^{-2}\hat{q}_n\hat{x}_n(\hat{x}_{n+1} - \hat{s}_n\hat{x}_n) \\ & -\gamma(1+\gamma\lambda)\sigma_u^{-2}\hat{q}_n^2\hat{x}_n^3(\hat{x}_{n+1} - \hat{s}_n\hat{x}_n) \\ & \cdot(\gamma\hat{q}_n\hat{x}_n^2 + \sigma_u^2)^{-1},\end{aligned}$$

$$\begin{aligned}\nu_{n+1} = {} & \gamma\lambda^2\hat{q}_n - \gamma\lambda(2+\gamma\lambda)\sigma_u^{-2}\hat{q}_n\hat{x}_n^2 \\ & +\gamma(1+\gamma\lambda)\sigma_u^{-2}\hat{q}_n^2\hat{x}_n^4(\gamma\hat{q}_n\hat{x}_n^2 + \sigma_u^2)^{-1}\end{aligned}$$

(notice that $\hat{x}'_n$ is obtained by the Kalman filter (2.15) with the covariance $\hat{p}_n$ replaced by $p(\hat{s}_n)$ of (3.1)). Furthermore, let $\xi_0 = (x_0, 0, \hat{x}_0)^T$ and

$\theta_n = (\hat{s}_n, \hat{q}_n)$, $\xi_{n+1} = (x_{n+1}, \hat{x}'_n, \hat{x}'_{n+1})^T$, $\rho_{n+1} = (\mu_{n+1}, \nu_{n+1})$, $\eta_n = (u_n, v_n)^T$. Then, the difference equations (1.3) and (2.13) – (2.16) can be represented in the following form:

$$s_{n+1} = s_n + \gamma\lambda s_n + \gamma w_{n+1}, \qquad (3.2)$$

$$\theta_{n+1} = \Pi_Q(\theta_n + \gamma H(\theta_n, \xi_{n+1}) + \gamma\rho_{n+1}), \qquad (3.3)$$

$$\xi_{n+1} = A(\theta_n, s_n)\xi_n + B(\theta_n)\eta_{n+1}, \qquad (3.4)$$

where $\Pi_Q(t) = \arg\min_{\theta' \in Q \times [0,k]} \|\theta - \theta'\|$, $\theta \in \mathbb{R}^2$. On the other hand, A1 and (3.4) imply that

$$\mathbb{P}(\xi_{n+1} \in B | \mathcal{F}_n) = P(B|\theta_n, s_n, \xi_n) \;\; w.p.1,$$

for all $B \in \mathcal{B}^3$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma\{u_i, v_i, w_i : 1 \le i \le n\}$, $n \ge 1$, and

$$P(B|\theta, s, \xi) = \mathbb{E}(I_B(A(\theta, s)\xi + B(\theta)\eta_0)), \quad B \in \mathcal{B}^3, \theta \in \mathbb{R}^2, s \in \mathbb{R} \qquad (3.5)$$

Hence, the difference equations (1.1) – (1.3) and (2.13) – (2.16) are really of the same form as stochastic approximation algorithms studied in (Benveniste *et al.*, 1990), and therefore, the asymptotic analysis of (1.1) – (1.3) and (2.13) – (2.16) is based on the methods developed in (Benveniste *et al.*, 1990). Notice that the above formulation decomposes naturally into two time scales: The 'hypermodel' $\{s_n\}_{n\ge0}$ in (3.2) is a slowly time varying process with small parameter $\gamma > 0$, while the regression vector ξ_n of (3.4) is rapidly time varying in comparison with s_n. The state estimation tracking algorithm (3.3) comprises of the Kalman filter state estimate $\hat{s}_n$ and covariance estimate $\hat{q}_n$ of the hypermodel $\{s_n\}_{n\ge0}$.

Throughout the paper, the following notation is used. For $\hat{s} \in \mathbb{R}, s \in Q$, let

$$r = \sigma_v^2(c^2\sigma_u^2 + \sigma_v^2)^{-1}, \quad \rho_Q = \sup_{t\in Q} \max\{r^{1/4}, |t|^{1/4}\},$$

$$\tilde{u}(\hat{s}, s) = \sigma_u^2\beta^2(\hat{s})(1 + s\alpha(\hat{s})) + c^{-2}\sigma_v^2\hat{s}\alpha(\hat{s})\beta^3(\hat{s})(1 - s^2),$$

$$\tilde{v}(\hat{s}, s) = \sigma_u^2\beta^2(\hat{s})(1 + s\alpha(\hat{s})) + c^{-2}\sigma_v^2\beta^2(\hat{s})(1 - s^2)(1 - s\alpha(\hat{s})),$$

$$\tilde{w}(\hat{s}, s) = (1 - s^2)(1 - s\alpha(\hat{s}))(1 - s^2),$$

$$a(\hat{s}, s) = \tilde{u}(\hat{s}, s)\tilde{w}^{-1}(\hat{s}, s),$$

$$b(\hat{s}, s) = \tilde{v}(\hat{s}, s)\tilde{w}^{-1}(\hat{s}, s),$$

$$\gamma(\hat{s}) = \sigma_v^2(c^2 p(\hat{s}) + \sigma_v^2)^{-1}$$

(obviously, $r, \rho_Q < 1$). Moreover, for $\hat{s} \in [-1, 1], \hat{q} \in [0, k], s \in Q, \theta = (\hat{s}, \hat{q})$, let

$$u(\hat{s}, s) = b^{-1}(\hat{s}, s)(\lambda\sigma_u^2 + (\lambda^2\sigma_u^4 + \sigma_u^2\sigma_w^2 b(\hat{s}, s))^{1/2}),$$

$$v(\hat{s}, \hat{q}, s) = \sigma_u^{-2}(-2\lambda\sigma_u^2 + b(\hat{s}, s)(\hat{q} + u(\hat{s}, s))),$$

$$f(\theta, s) = \lambda\hat{s} - \hat{q}(\hat{s} - s)a(\hat{s}, s),$$

$$g(\theta, s) = -(\hat{q} - u(\hat{s}, s))a(\hat{s}, s)$$

and $w(s) = u(s, s)$.

4. MAIN RESULTS

In order to present the main results of this paper, the following notation is needed. Let $\Theta_n = (\theta_n, s_n)$. Moreover, for $\theta \in [-1, 1] \times [0, k], s \in Q, \Theta = (\theta, s)$, let

$$\xi_n(\theta, s) = \sum_{i=n}^{\infty} A^{i-n}(\theta, s)B(\theta)\eta_i, \qquad (4.1)$$

$$S(\theta, s) = \sum_{n=0}^{\infty} \operatorname{cov}(H(\theta, \xi_0(\theta, s)), H(\theta, \xi_n(\theta, s))) + \sum_{n=1}^{\infty} \operatorname{cov}^T(H(\theta, \xi_0(\theta, s)), H(\theta, \xi_n(\theta, s))),$$

while $h(\theta, s) = \mathbb{E}(H(\theta, \xi_0(\theta, s)))$, $R(\Theta) = (h(\theta, s), \lambda s)$, $\tilde{R}(\Theta) = \partial R(\Theta)$, $\tilde{S}(\Theta) = \operatorname{diag}\{S(\theta, s), \sigma_w^2\}$. Moreover, let $\Theta(\cdot)$ and $\tilde{\Theta}(\cdot)$ be the solutions of the equations

$$\Theta(t) = \Theta_0 + \int_0^t R(\Theta(\tau))d\tau, \quad t \in \mathbb{R}_0^+, \qquad (4.2)$$

$$\tilde{\Theta}(t) = \int_0^t \tilde{R}(\tilde{\Theta}(\tau))d\tau + \int_0^t \tilde{S}(\tilde{\Theta}(\tau))\tilde{W}(d\tau), \quad t \in \mathbb{R}_0^+, \qquad (4.3)$$

where $\tilde{W}(\cdot)$ is standard Brownian motion, while $\Theta(t) = (\hat{s}(t), \hat{q}(t), s(t))$, $\tilde{\Theta}(t) = (\tilde{s}(t), \tilde{q}(t), \tilde{s}'(t))$ and $\tilde{e}(t) = \tilde{s}(t) - \tilde{s}'(t)$ for all $t \in \mathbb{R}_0^+$. Furthermore, for $t \in \mathbb{R}_0^+, \gamma \in \mathbb{R}^+$, let

$$\Theta_\gamma(t) = \sum_{n=0}^{\infty} (\Theta_n + \gamma^{-1}(t - n\gamma)(\Theta_{n+1} - \Theta_n)) \cdot I_{[n\gamma,(n+1)\gamma)}(t)$$

and $\tilde{\Theta}_\gamma(t) = \gamma^{-1/2}(\Theta_\gamma(t) - \Theta(t))$, $\Theta_\gamma(t) = (\hat{s}_\gamma(t), \hat{q}_\gamma(t), s_\gamma(t))$, $\tilde{\Theta}_\gamma(t) = (\tilde{s}_\gamma(t), \tilde{q}_\gamma(t), \tilde{s}'_\gamma(t))$ and $\tilde{e}_\gamma(t) = \tilde{s}_\gamma(t) - \tilde{s}'_\gamma(t)$.

Remark. $h(\cdot, \cdot)$ and $S(\cdot, \cdot)$ can be considered as the 'averaged function' and 'limiting covariance' associated with the difference equations (3.3), (3.4). Similarly, $R(\cdot)$ and $\tilde{S}(\cdot)$ are the 'averaged function' and 'limiting covariance' associated with the difference equations (3.2) – (3.4). On the other hand, (4.2) and (4.3) can be considered as the 'averaged' ordinary differential equation (ODE) and 'limiting' stochastic differential equation (SDE) associated with (3.2) – (3.4). The asymptotic behavior of $\{\Theta_n\}_{n\ge0}$ as $n \to \infty$ and $\gamma \to 0+$ is tightly connected to (4.2) and (4.3). Namely, the piecewise linear interpolation $\Theta_\gamma(\cdot)$ of $\{\Theta_n\}_{n\ge0}$ converges to a solution of (4.2), while the piecewise linear interpolation $\tilde{\Theta}_\gamma(\cdot)$ of $\{\gamma^{1/2}(\Theta_n - \Theta(n\gamma))\}_{n\ge0}$ converges weakly (in distribution) to a solution of of the SDE (4.3) (this can be viewed as a functional central limit theorem; for more details on the relevance of the associated ODE and SDE for the asymptotic analysis of stochastic approximation algorithms see (Benveniste *et al.*, 1990)).

The main results of the paper are contained in the following theorem:

Theorem 1. Let A1 hold. Then, $\{\tilde{\Theta}_\gamma(\cdot)\}_{\gamma\in\mathbb{R}^+}$ converges weakly in Skorohod topology to $\tilde{\Theta}(\cdot)$ as $\gamma \to 0+$.

For the proof see (Tadić and Krishnamurthy, 2001).

As immediate consequences of Theorem 1, the following corollary is obtained:

Corollary 1. Let A1 hold. Then,

$$\hat{s}_\gamma(t) - s_\gamma(t) = \hat{s}(t) - s(t) + \gamma^{1/2}\tilde{e}_\gamma(t) \qquad (4.4)$$

for all $t \in \mathbb{R}_0^+, \gamma \in \mathbb{R}^+$. Moreover, $\{\tilde{e}_\gamma(\cdot)\}_{\gamma\in\mathbb{R}^+}$ converges weakly in Skorohod topology to $\tilde{e}(\cdot)$ as $\gamma \to 0+$.

Remark. The results of Theorem 1 are related to the weak convergence of $\{\Theta_n\}_{n\geq 0}$ as $\gamma \to 0+$, while Corollary 1 corresponds with the weak convergence of the error of the estimates $\hat{s}_n$ as $\gamma \to 0+$. Moreover, Corollary 1 provides an insight into the structure of this error — (4.4) represents the decomposition of the error into the sum of a bias and variance, where $\hat{s}(\cdot) - s(\cdot)$ is the bias and $\gamma^{1/2}\tilde{e}_\gamma(\cdot)$ is the variance.

5. CONCLUSIONS

We have presented an asymptotic analysis of the cross-coupled Kalman filter algorithm for estimating the state of a scalar valued partially observed bilinear system. Numerical studies given in (Johnston and Krishnamurthy, 1999) shown that the cross-coupled Kalman filter algorithm performs significantly better than the Extended Kalman filter in several situations. In future work we will extend the analysis to jump Markov linear systems – where the estimation algorithm consists of cross-coupling a Kalman filter with a Hidden Markov Model filter.

6. REFERENCES

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* Prentice-Hall.

Benedetto, S. and E. Biglieri (1983). Nonlinear equalization of digital satellite channels. *IEEE Journal of Selected Areas in Communications* **1**, 57–62.

Benveniste, A., M. Metivier and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximation.* Springer Verlag.

Bruni, C., G. DiPillo, and G. Koch (1974). Bilinear systems: an appealing class of 'nearly linear systems' in theory and applications. *IEEE Transactions on Automatic Control* **19**, 334–348.

Caines, P. E. (1988). *Linear Stochastic Systems.* Wiley.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* **39**, 1–38.

Fnaiech, F. and L. Ljung (1987). Recursive identification of bilinear systems. **45**, 453–470.

Halawani, T.U., R.R. Mohler and W.J. Kolodziej (1984). A two-step bilinear filtering algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **32**, 344–352.

Johnston, L. and V. Krishnamurthy (1999). Finite dimensional smoothers for MAP state estimation of bilinear systems. *IEEE Transactions on Signal Processing* **47**, 2444–2459.

Ljung, L. (1979). Asymptotic behaviour of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions Automatic Control* **24**, 36–50.

Luenberger, D.G. (1984). *Linear and Nonlinear Programming.* Addison Wesley.

Priestley, M.B. (1991). *Non-linear and Non-stationary Time Series Analysis.* Academic Press.

Tadić, V. B. and V. Krishnamurthy (2001). Asymptotic analysis of cross-coupled kalman filter state-estimation algorithm for bilinear systems. Technical report. Department of Electrical and Electronic Enginering, The University of Melbourne.

Wu, C.F.J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics* **11**, 95–103.